Geometric Distribution

Characteristics of a Geometric distribution:

1. Each observation (or trial) has two categories: success or failure.
2. The observations are all independent.
3. The probability of success (p) is the same for each trial.
4. X be the number of failures before the first success **OR** X be the number of trials needed for the first success.
5. We wish to find the number of trials needed to obtain the first success.

What are the Differences between the Geometric and the Binomial Distributions?

1. The most obvious difference is that the Geometric Distribution does not have a set number of observations, n.
2. The second most obvious difference is the question being asked:
   Binomial: Asks for the probability of a certain number of successes.
   Geometric: Asks for the probability of the first success.

Suppose a coin is tossed till a head (H) occurs for the first time we may get a head at the 1st trial 2nd trial or at the 3rd trial and so on let X denote the number of failures before getting a head for the first time we want to find probability distribution of the discrete random variable X suppose , the probability of getting head in a single trial is p , probability of getting tail in a single trial ,
q =1-p $\quad [\because p+q=1]$
Assuming that the trials are independent , the probability distribution of X will be as follows :

X	0	1	2	3	….	X	….
P(X)	P	qp	q^2p	q^3p	……	q^xp	…..

Such a probability distribution is called geometric distribution with parameter P . since, these probabilities form a geometric progression.

Hence , $P\{X=x\}=q^xp$; $\quad$ X=0,1,2,3….

Geometric distribution is a applied in the field of reliability and queueing theory (see real life situation

Probability mass function of geometric distribution

Consider a sequence of Bernoulli trials, with consist probability of success p and q be the failures probability. Let X represent the number of trials needed for the first success.

$$S = \{S, FS, FFS, FFFS, FFFFS, \ldots\ldots\ldots..\}$$

Hence, $P\{X = x\} = q^{x-1}p; \quad X=1,2,3\ldots.$

Definition:

A random variable X is said to be follow a geometric distribution if its probability mass function is given by

$P(X=x) = q^{x-1}p \quad ; \quad x=1,2,3,...$

$P(X=x) = 0 \quad ; \quad \text{otherwise}$

$0 < p < 1 \ ; \ p + q = 1$

Here, X be the number of trials needed for the first success.

OR

Definition:

A random variable Y is said to be follow a geometric distribution if its probability mass function is given by

$P(Y= y) = q^{x-1}p \quad ; \quad y=1,2,3,...$

$P(Y=y) = 0 \quad ; \quad \text{otherwise}$

$0 < p < 1 \ ; \ p + q = 1$

Here, Y be the number of failures before the first success.

Remark:

[1] $\left[1+x+x^2+....+x^n\right] = [1-x]^{-1} = \dfrac{1}{1-x}$

[2] $\left[1+2x+3x^2+...\right] = (1-x)^{-2}$

Now, the mean of the Geometric distribution is

$$E(X)=\sum_{x=0}^{\infty} x\,P(X=x)$$

$$E(X)=\sum_{x=0}^{\infty} x\,pq^{x}$$

$$E(X)=0\,pq^{0}+1pq^{1}+2\,pq^{2}+3\,pq^{3}+\ldots$$

$$E(X)=pq\left[1+2q+3q^{2}+\ldots\right]$$

$$E(X)=pq\left[1-q\right]^{-2} \qquad \because \left[1+2x+3x^{2}+\ldots\right]=(1-x)^{-2}$$

$$E(X)=\frac{pq}{\left[p\right]^{2}} \qquad \because 1-q=p$$

$$E(X)=\frac{q}{p}$$

$\therefore$ The mean of the Geometric distribution is $\dfrac{q}{p}$

- **Variance of the Geometric distribution:**

$$V(X)=E(X^{2})-[E(X)]^{2}$$

$$V(X)=E(X^{2})-\left[\frac{q}{p}\right]^{2} \qquad \ldots(i) \quad \left(\because E(X)=\frac{q}{p}\right)$$

Now,

$$E(X^{2})=\sum_{x=0}^{\infty} x^{2}\,P(X=x)$$

$$E(X^{2})=\sum_{x=0}^{\infty} x\left(x-1+1\right)pq^{x}$$

$$E(X^{2})=\sum_{x=0}^{\infty} x\left(x-1\right)pq^{x}+\sum_{x=0}^{\infty} x\,pq^{x}$$

$$E(X^{2})=\sum_{x=0}^{\infty} x\left(x-1\right)pq^{x}+\frac{q}{p}$$

$$E(X^{2})=2\,pq^{2}+6\,pq^{3}+12\,pq^{4}+\ldots+\frac{q}{p}$$

$$E(X^{2})=2\,pq^{2}\left[1+3q+6q^{2}+\ldots\right]+\frac{q}{p}$$

$$2pq^{2}\left[1-q\right]^{-3}+\frac{q}{p} \qquad \because \left[1+2x+3x^{2}+\ldots\right]=(1-x)^{-2}$$

$$E(X^2) = \frac{2pq^2}{[p]^3} + \frac{q}{p} \qquad Q\ 1-q = p$$

$$E(X^2) = \frac{2q^2}{p^2} + \frac{q}{p} \qquad \ldots(ii)$$

Putting (ii) in (i), we get

$$V(X) = \frac{2q^2}{p^2} + \frac{q}{p} - \frac{q^2}{p^2}$$

$$V(X) = \frac{q^2}{p^2} + \frac{q}{p}$$

$$V(X) = \frac{q}{p}\left[\frac{q}{p} + 1\right]$$

$$V(X) = \frac{q}{p}\left[\frac{q+p}{p}\right]$$

$$V(X) = \frac{q}{p}\left[\frac{1}{p}\right] \qquad \because q+p = 1$$

$$V(X) = \frac{q}{p^2}$$

$\therefore$ The variance of the Geometric distribution is $\frac{q}{p^2}$

Moment generating function (M.G.F.) of G(p):

$$M_X(t) = E[e^{tx}]$$

$$M_X(t) = \sum_{x=0}^{\infty} e^{tx}\, p(x)$$

$$M_X(t) = \sum_{x=0}^{\infty} e^{tx} q^x . p = p \sum_{x=0}^{\infty} \left(qe^t\right)^x$$

$$M_X(t) = P\left(1 + (qe^t) + (qe^t)^2 + (qe^t)^3 + \ldots\ldots..\right) \quad (\text{if } \left|qe^t\right| < 1 \quad \text{i.e. for } t < -\log q)$$

$$M_X(t) = P\left(1 - qe^t\right)^{-1} \qquad \left(\because 1 + x + x^2 + x^3 + x^4 + \cdots = (1-x)^{-1}\right)$$

$$M_X(t) = \frac{p}{1 - qe^t}$$

Cumulant generating function [C.G.F.] of Geometric Distribution:

We know that the m.g.f. is given by,

$$M_X(t)=\frac{p}{1-qe^t};t<-\log q$$

$$M_X(t)=\frac{p}{p+q-qe^t}=\frac{1}{1+\frac{q}{p}-\frac{q}{p}e^t}$$

$$M_X(t)=\frac{1}{1-\frac{q}{p}(e^t-1)}$$

Therefore, the cumulant generating function will be,

$$K_X(t)=\log_e\left[M_X(t)\right] \qquad (\because \text{By definition of C.G.F.})$$

$$K_X(t)=\log_e\left\{\frac{1}{\left[1-\frac{q}{p}(e^t-1)\right]}\right\}$$

$$K_X(t)=\log\left[1-\frac{q}{p}(e^t-1)\right]^{-1}$$

$$K_X(t)=-\log\left[1-\frac{q}{p}(e^t-1)\right] \qquad (\because \text{log of exponant})$$

$$\therefore K_X(t)=-\log\left[1-\frac{q}{p}(e^t-1)\right]$$

Deduction:

$$K_X(t)=-\log\left[1-\frac{q}{p}(e^t-1)\right]$$

$$K_X(t)=-\log\left[1-\frac{q}{p}\left(1+t+\frac{t^2}{2!}+\frac{t^3}{3!}+\cdots-1\right)\right] \quad \left(\because e^t=\left(1+t+\frac{t^2}{2!}+\frac{t^3}{3!}...\right)\right)$$

using $-\log(1-x)=\left(x+\frac{x^2}{2}+\frac{x^3}{3}+...\right)$ we have,

$$K_X(t)=\frac{q}{p}\left(t+\frac{t^2}{2!}+\frac{t^3}{3!}+...\right)+\frac{1}{2}\left(\frac{q}{p}(t+\frac{t^2}{2!}+\frac{t^3}{3!}+...)\right)^2$$

$$+\frac{1}{3}\left(\frac{q}{p}(t+\frac{t^2}{2!}+\frac{t^3}{3!}+...)\right)^3+\frac{1}{4}\left(\frac{q}{p}(t+\frac{t^2}{2!}+\frac{t^3}{3!}+...)\right)^4+\cdots$$

and we have, $K_X(t) = K_1 t + K_2 \frac{t^2}{2!} + K_3 \frac{t^3}{3!} + K_4 \frac{t^4}{4!} + \cdots$

comparing coefficient of $\frac{t^r}{r!}$ we get

$$K_1 = \frac{q}{p}, K_2 = \frac{q}{p^2}, K_3 = \frac{q(1+q)}{p^3}, K_4 = \frac{q(p^2+6q)}{p^4}$$

$$\text{Hence } \mu_1 = \frac{q}{p}, \mu_2 = \frac{q}{p^2}, \mu_3 = \frac{q(1+q)}{p^3}, \mu_4 = \frac{q}{p^4}(p^2+6q)$$

Recurrence Relation between Probabilities of Geometric Distribution

We establish the relation between P(X= x) and P(X= x+1).

Note that:

$$P(X=x) = q^x p; \qquad 0< p <1: \qquad X=0,1,2,3........$$

$$P(X=x+1) = pq^{x+1}$$

$$\therefore \frac{P(X=x+1)}{P(X=x)} = q$$

$$\therefore P(X=x+1) = q \times p(X=x)$$

This is the recurrence relation between probabilities of geometric distribution. If we know P(X=0) = p, then we can obtain P(X=1), P(X=2) and so on using above relationship.

Note: The same recurrence relation holds good for another form of p.m. f. of geometric distribution.

Distribution function of geometric Distribution

$$F_y(y) = P(Y \leq y) = \sum_{r=1}^{y} q^{r-1} p$$

$$F_y(y) = P(1+q+q^2+...+q^{y-1})$$

The bracket contains sum of first y terms of G.P. with first term 1 and common ratio q.

$$F_y(y) = p.\frac{1-q^y}{1-q}$$

$$F_y(y) = p.\frac{1-q^y}{p}$$

$$F_y(y) = 1-q^y; \qquad y=1,2,3,\ldots$$

Memory less property of geometric distribution

Statement :

X is non- negative R. V., it has p.m.f. $P\left(\frac{x>t+s}{X>t}\right)=P(X>S);\forall s,t\geq 0$ (s and t are positive integers).

Proof: $X\sim Geo(P)$, where X be the no of trials needed for the first success.

$F_y(y)=P(Y\leq y)=1\text{-}q^y=1\text{-}q^y$, q+p =1

Consider, L.H.S = $P\left(\frac{x>t+s}{X>t}\right)=P\left(\frac{x>t+s\cap x>t}{X>t}\right)$

Now,

$$P\left(\frac{x>t+s}{X>t}\right)=P(X>t+s)$$

$$P\left(\frac{x>t+s}{X>t}\right)=1\text{-}P(X\leq t+s)$$

$$P\left(\frac{x>t+s}{X>t}\right)=1\text{-}\left[1\text{-}q^{t+s}\right] \qquad (\because \text{By definition of distribution function})$$

$$P\left(\frac{x>t+s}{X>t}\right)=q^{t+s}\ldots\ldots(1)$$

And

$$P(X>t)=1\text{-}P(X\leq t)$$

$$P(X>t)=1\text{-}\left[1\text{-}q^t\right] \qquad (\because \text{By definition of distribution function})$$

$$P(X>t)=q^t\ldots\ldots(2)$$

L.H.S. :

$$P(X>t)=\frac{q^{t+s}}{q^t}$$

$$P(X>t)=q^s\ldots\ldots(3)$$

R.H.S. :

$$P(X>s)=1\text{-}P(X\leq s)$$

$$P(X>s)=1\text{-}\left[1\text{-}q^s\right]$$

$$P(X>s)=q^s \qquad \ldots\ldots(4)$$

From (3) & (4), We get

L.H.S. = R.H.S

$$P\left(\frac{x > t+s}{X > t}\right) = P(X > S); \quad \forall s, t \geq 0$$

Example:

Suppose an electronic component or fuse fail to work at x^{th} hour for the first time. Then $P\left(\frac{x > t+s}{X > t}\right)$ is probability that a component not failed in s hours will also not fail up to (s + t) hours. It means the component will not fail for next t hours. It means the component will not fail in next hours is irrespective of the no. of hours it is not failed in past. Thus, it forgets its past working time and works like a brand new component. Therefore, the properties described as memoryless property or forgetfulness property.

Que. 1 Fill in the blanks and complete the following statements:

[1] If Bernoulli trials are condected till setting the first success then the probability distribution of number of

a. Trials is………

b. Failure before the first success is…….

[2] If $P(x) = \frac{1}{2^x}$; x= 1,2,…… then,

a.E(X)=…..

b. Var(X)=…..

c. Mode of X is…..

[3] If X is the geometric random variable then p(X> 5/X>2)=…..

[4] The variance of geometric distribution is ……….than the mean .

[5] The M.G.F. of geometric distribution on a set of

a. non-negative integers is …..

b. positive integers is …….

[6] . The mean of geometric distribution on a set of

a. non negative integers is…….

b. positive integers is…….

[7] The variance of geometric distribution is …….

[8] If X is discrete r.v. satisfying lack of memory property then the probability distribution of X is …..

[9] If $X \rightarrow$ geometric (p) then p(X=0) , p(X=1), ……. From a…… progression

[10] If $X \rightarrow$ geometric (p) then $\frac{p(X = x+1)}{p(X = x)}$ =……..

[11] If $X \rightarrow$ geometric (p) then the ratio of probabilities taking any two successive values is……

(B). choose the correct alternative (Q.12 to Q.14):

[12] If $X \sim>$ geometric (p), then ……

a. p(X> s+t/x>s)=p(x>s) b. p(x>s+t/x>s)=p(x>s+t)

c. p(X>s+t/x>s)=p(x>t) d. p(X>s+t/x>s)=p(x<t)

[13] If $X \sim>$ geometric (p) taking values 1,2……… then

a. mean > variance b. mean < variance

c. mean = variance d. mean = 2 variance

[14] If X= number of candidates required to interview for the post of an officer then X follows:

a. binomial distribution b. Bernoulli distribution

c. poisson distribution d. geometric distribution